SF 298 last page

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June, 1998	Technical Report; June 1998

4. TITLE AND SUBTITLE
On the Performance of Subset Selection Procedures Under Normality

5. FUNDING NUMBERS
DAAH04-95-1-0165

6. AUTHOR(S)
Shanti S. Gupta and Klaus J. Miescke

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Purdue University
Department of Statistics
West Lafayette, IN 47907-1399

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report #98-18

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 32922.17-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

From k normal populations $N(\theta_1, \sigma_1^2), \ldots, N(\theta_k, \sigma_k^2)$, where the means $\theta_1, \ldots, \theta_k \in \Re$ are unknown, and the variances $\sigma_1^2, \ldots, \sigma_k^2 > 0$ are known, independent random samples of sizes $n_1, \ldots, n_k$, respectively, are drawn. Based on these observations, a non-empty subset of these k populations of preferably small size has to be selected, which contains the population with the largest mean with probability of the lest P^* at every parameter configuration. Several subset selection procedures which have been proposed in the literature are compared with Bayes selection procedures for normal priors under two natural type of loss functions. Two new subset selection procedures are considered.

14. SUBJECT TERMS
Subset selection procedures; Bayes selection procedures; admissible subset selection rules; normal populations

15. NUMBER IF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ON THE PERFORMANCE OF SUBSET SELECTION PROCEDURES UNDER NORMALITY

by

Shanti S. Gupta
Purdue University

and

Klause J. Miescke
University of Illinois
at Chicago

Technical Report # 98-18

Department of Statistics
Purdue University

June 1998

△

ON THE PERFORMANCE OF SUBSET SELECTION PROCEDURES UNDER NORMALITY *

Shanti S. Gupta
Department of Statistics
Purdue University
Math. Sciences Bldg.
West Lafayette, IN 47907

Klaus J. Miescke
Department of MSCS (M/C 249)
University of Illinois at Chicago
851 South Morgan St.
Chicago, IL 60607-7045, U.S.A.

ABSTRACT. From k normal populations $N(\theta_1,\sigma_1^2),\ldots,N(\theta_k,\sigma_k^2)$, where the means $\theta_1,\ldots,\theta_k \in \Re$ are unknown, and the variances $\sigma_1^2,\ldots,\sigma_k^2 > 0$ are known, independent random samples of sizes $n_1,\ldots,n_k$, respectively, are drawn. Based on these observations, a non-empty subset of these k populations of preferably small size has to be selected, which contains the population with the largest mean with a probability of at least P^* at every parameter configuration. Several subset selection procedures which have been proposed in the literature are compared with Bayes selection procedures for normal priors under two natural type of loss functions. Two new subset selection procedures are considered.

AMS(1980) Subject Classification. Primary 62F07; Secondary 62F15.

Key Words and Phrases. Subset selection procedures; Bayes selection procedures; admissible subset selection procedures; independent normal populations.

---

* This research was supported in part by US Army Research Office Grant DAAH04-95-1-0165 at Purdue University.

1. INTRODUCTION

Let $P_1,...,P_k$ be k normal populations $N(\theta_1,\sigma_1^2),...,N(\theta_k,\sigma_k^2)$, respectively, where the means $\theta_1,...,\theta_k \in \Re$ are unknown, and where the variances $\sigma_1^2,...,\sigma_k^2 > 0$ are known. Suppose one wants to select, based on independent random samples of respective sizes $n_1,...,n_k$, a non-empty subset of these populations which contains the population with the largest θ-value, i.e. the best population, with a high probability, where the size of the subset should be as small as possible. Apparently both goals work against each other, similar as in testing hypotheses where the goal is to keep both types of error probability small. As in the latter, a natural approach is to control one of the two goals and then to optimize the other one.

The classical approach, due to Gupta (1956,65), is to consider subset selection procedures S for which the minimum probability of a correct selection (CS) is at least P^*, where P^* is predetermined. To exclude the possibility of simply selecting just one of the populations at random, P^* is assumed to be greater than k^{-1}. Among all subset selection procedure S which satisfy this P^*-condition, an optimum procedure would have the smallest expected size $E_{\underline{\theta}}(|S|)$ for all $\underline{\theta} \in \Re^k$. Unfortunately, no such optimum procedure exists (Deely and Johnson, 1997). Therefore, attempts have been made in the past to find suitable candidates which perform well in terms of their expected subset sizes in comparison to the other procedures.

By sufficiency, only subset selection procedures have to be considered which depend on the k sample means $\underline{X} = (X_1,...,X_k)$ with realizations $\underline{x} = (x_1,...,x_k) \in \Re^k$. The underlying model simplifies to $X_i \sim N(\theta_i, p_i^{-1})$, where $p_i = n_i / \sigma_i^2$ denotes the precision of the sample mean from population P_i, $i = 1,...,k$, and $X_1,...,X_k$ are independent. For notational convenience, let

in the sequel $\vartheta_{[1]} \le \vartheta_{[2]} \le ... \le \vartheta_{[k]}$ denote the ordered values of any $\underline{\vartheta} \in \Re^k$. Further, let φ and Φ denote the density and the c.d.f., respectively, of the standard normal distribution $N(0,1)$.

For the case of $\sigma_1^2 = ... = \sigma_k^2 = \sigma^2 > 0$ and $n_1 = ... = n_k = n_0$, Gupta (1956,65) has proposed the following, now classical, subset selection procedure.

$$S_{Gupta}(\underline{X}) = \left\{ i \,\middle|\, X_i \ge X_{[k]} - d\sigma/\sqrt{n_0}, i = 1,...,k \right\}, \text{ where } \int_{\Re} \Phi^{k-1}(z+d)\,\varphi(z)\,dz = P^*. \qquad (1)$$

Modifications of this procedure to cases where not both, the variances and the sample sizes, are common, have been proposed by Chen and Dudewicz (1973), Gupta and Huang, W.T. (1974), Gupta and Huang, D.Y. (1976), and Gupta and Wong (1976). They are all of the common form $S_G(\underline{X}) = \left\{ i \,\middle|\, X_i \ge X_j - c_{i,j}\ j \ne i,\ i = 1,...,k \right\}$, where the $c_{i,j}$'s are positive, and are chosen to meet the P^*-condition in their given settings. Berger and Gupta (1980) have shown that S_{Gupta} and some, but not all, of its modifications of the type S_G to other settings are minimax, in terms of the expected subset size, in the class of subset selection procedures which satisfy the P^*-condition, and are non-randomized, just, and translation invariant. Gupta and Miescke (1981) have shown that for $k = 3$, S_G is optimum, in terms of the expected sample size, in the class of subset selection procedures proposed by Seal (1955,57), when the distances between $\theta_1,...,\theta_k$ are larger than certain positive constants. Similar results for a particular slippage configuration of the parameters had been derived previously by Deely and Gupta (1968). Procedures of this type have been the objects of numerous investigations in the literature in the past. An overview and further references can be found in the monographs by Gupta and Panchapakesan (1979) and by Bechhofer, Santner, and Goldsman (1995), and recent results in Miescke and Rasch (1996).

The purpose of the present study is to look at the above mentioned subset selection procedures from a decision theoretic point of view, by means of comparisons with Bayes

selection procedures under two natural type of loss function, one of which being closely related to the problem of minimizing the expected sample size subject to the P^*-condition.

2. Losses and Frequentist Risks

Let $L(\underline{\theta},s)$ be a given loss function for selecting the populations P_i with $i \in s$ at $\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Re^k$. A non-randomized subset selection procedure S is a measurable function from the sampling space $\Re^k$ into the set $\left\{s \middle| \varnothing \neq s \subseteq \{1,\ldots,k\}\right\}$, where the elements of S are the indices of the populations that are selected. More generally, one could consider also randomized subset selection procedures S^*, where at every $\underline{X} = \underline{x}$, population P_i is selected by S^* with probability $p_{S^*,i}(\underline{x})$, $i = 1,\ldots,k$. However, since this does not lead to any further improvement, such considerations will be postponed until the end.

Consider now a natural loss function, with some variations, that supports the goals given in Section 1. One possible conflict, however, should be discussed first. It is not immediately clear if situations where populations are tied for the largest parameter are statistically relevant or not. One could argue that nothing in real life repeats itself in an identical manner. On the other hand, two of the populations could indeed be associated with the same experimental setup. For the sake of completeness, suppose that the occurrence of such ties is statistically relevant. Then several variations of the natural loss function have to be distinguished.

$$
\begin{aligned}
L_1(\underline{\theta},s) &= \sum_{i\in s}\left[\, a - I_{\{\theta_{[k]}\}}(\theta_i) \,\right], \qquad (2)\\
L_2(\underline{\theta},s) &= \sum_{i\in s}\left[\, a - I_{\{\tau(\underline{\theta})\}}(i) \,\right], \quad \tau(\underline{\theta}) \in \{i \mid \theta_i = \theta_{[k]}\},\\
L_3(\underline{\theta},s) &= \sum_{i\in s} a - I_{\{\theta_{[k]}\}}(\max_{i\in s}\{\theta_i\}), \quad \underline{\theta} \in \Re^k, \quad \varnothing \neq s \subseteq \{1,\ldots,k\}, \; 0 < a < 1.
\end{aligned}
$$

The first version, L_1, rewards inclusion of all populations which have the largest parameter, the second, L_2, rewards inclusion of exactly one population that has been *tagged* as the best by a given function τ, and the third, L_3, rewards the inclusion of only one, but any one, population

which has the largest parameter. The choice of $a = 0$ $(a \geq 1)$ has been omitted since it supports the selection of all k (only 1 of the) populations. A positive factor b could have been attached to the indicator functions in (2), with the restriction of $0 < a < b$, but without gain of generality.

A modified version of L_1, where the gain of including a best population depends on the number of populations tied for the best, has been used in Gupta and Hsu (1977). L_2 has been used only implicitly in the literature, as it can be understood from the respective contexts. L_3 has been used in Bratcher and Bhalla (1974) and in simulation studies of Gupta and Hsu (1978). In the sequel, situations where populations are tied for the best will have probability zero, and then distinctions between L_1, L_2, and L_3 will be irrelevant. L_1 proves then to be the most convenient version, and will thus be used. Let it be called henceforth *"(0,a,a-1) loss"*, and denoted by L.

The frequentist risk of a subset selection procedure S under loss L is given by

$$R(\underline{\theta},S) = E_{\underline{\theta}}\left(L(\underline{\theta},S(\underline{X}))\right) = \sum_{i=1}^{k}\left[a - I_{\{\theta_{[k]}\}}(\theta_i)\right] P_{\underline{\theta}}\{i \in S(\underline{X})\} \qquad (3)$$
$$= a\,E_{\underline{\theta}}\left(\left|S(\underline{X})\right|\right) - \sum_{i \in A(\underline{\theta})} P_{\underline{\theta}}\{i \in S(\underline{X})\},$$

where $A(\underline{\theta}) = \{j \mid \theta_j = \theta_{[k]}\}$, $\underline{\theta} \in \Re^k$. On $\Omega = \{\underline{\theta} \mid \theta_{[k-1]} < \theta_{[k]}, \underline{\theta} \in \Re^k\}$, this simplifies to

$$R(\underline{\theta},S) = a\,E_{\underline{\theta}}\left(\left|S(\underline{X})\right|\right) - P_{\underline{\theta}}\{i^*(\underline{\theta}) \in S(\underline{X})\}, \text{ where } \theta_{i^*(\underline{\theta})} = \theta_{[k]},\ \underline{\theta} \in \Omega. \qquad (4)$$

The risk function of a subset selection rule S is not continuous on $\Re^k$, since in (3) the size of $A(\underline{\theta})$ can drop down to 1 in every neighborhood of a point $\underline{\theta} \in \Re^k$ with $\theta_{[k-1]} = \theta_{[k]}$. Obviously, for every subset selection procedures S, its minimum probability of a correct selection (CS) must occur on Ω. This fact will be used later on in Section 3.

Some interesting features of this risk will now be discussed. Consider the following scenarios, which may hold for any two given selection rules S_1 and S_2.

$$
\begin{aligned}
&\text{(a)} && R(\underline{\theta},S_1) \le R(\underline{\theta},S_2), && \underline{\theta} \in \Omega, \\
&\text{(b)} && P_{\underline{\theta}}\{\text{CS under } S_1(\underline{X})\} \ge P_{\underline{\theta}}\{\text{CS under } S_2(\underline{X})\}, && \underline{\theta} \in \Omega, \\
&\text{(c)} && E_{\underline{\theta}}(|S_1(\underline{X})|) \le E_{\underline{\theta}}(|S_2(\underline{X})|), && \underline{\theta} \in \Omega.
\end{aligned} \tag{5}
$$

Also, let (b') be (b) with "$\ge$" replaced by "$\le$ ", and (c') be (c) with "$\le$ " replaced by "$\ge$". Then (b) and (c) together imply (a), (a) and (b') together imply (c), and (a) and (c') together imply (b). Moreover, for a subset selection rule S_1 which is admissible, there cannot be a subset selection rule S_2 which satisfies both, (b') and (c'), with at least one strict inequality in (b') or (c'). In other words, one can state the following.

Lemma 1. *Let* Ω *be the parameter space, and* S *be a subset selection rule which is admissible in the class of non-randomized subset selection rules under loss* L. *Then there does not exist any subset selection rule which is as good as* S *in terms of both, probability of correct selection and expected subset size, and better in at least one of the two at some point in* Ω.

Since there does not exist a subset selection rule which minimizes the expected subset size, uniformly in $\underline{\theta} \in \Omega$, subject to the P^*-condition, it is reasonable to consider subset selection rules which are admissible and satisfy the P^*-condition. As it will be seen later, such rules can be found in the class of Bayes subset selection rules.

Other loss functions for subset selection procedures which have been used in the literature, depend more smoothly on the distances of the selected parameters to the largest parameter, and are discussed in Miescke (1997). A natural loss function in this class is the so-called *"linear loss"*, which has been used in Miescke (1979), and is given by

$$
L^*(\underline{\theta},s) = \sum_{i \in s}\left[\ \theta_{[k]} - \theta_i - \Delta\ \right], \quad \underline{\theta} \in \Re^k, \quad \varnothing \ne s \subseteq \{1,\ldots,k\}, \quad 0 < \Delta. \tag{6}
$$

Although not a linear function of $\underline{\theta}$, it increases linearly when non-largest parameters of populations in the selected subset are moving away from the largest, hence its name. More

specifically, under this loss function, selection of a population with its parameter less (more) than Δ away from the largest parameter is rewarded (penalized) on a linear scale.

The frequentist risk of a subset selection procedure S under loss L^* is given by

$$R^*(\underline{\theta},S) = E_{\underline{\theta}}\left(L^*(\underline{\theta},S(\underline{X}))\right) = \sum_{i=1}^{k}\left[\theta_{[k]} - \theta_i - \Delta\right] P_{\underline{\theta}}\left\{i \in S(\underline{X})\right\} \qquad (7)$$

$$= \left[\theta_{[k]} + \gamma - \Delta\right] E_{\underline{\theta}}(S|\underline{X}) - \sum_{i=1}^{k} [\theta_i + \gamma]\, P_{\underline{\theta}}\left\{i \in S(\underline{X})\right\}, \quad \underline{\theta} \in \Re^k,$$

where γ is chosen to have $\theta_{[k]} + \gamma - \Delta > 0$. Situations where populations are tied for the best do not cause the discontinuity conflicts that were observed under the *"(0,a,a-1) loss"*. Implications, analogously to those below of (5), hold for the terms in the lower part of (7), by simply replacing $P_{\underline{\theta}}\{\text{CS under } S(\underline{X})\}$ in (5) by $\sum_{i=1}^{k}[\theta_i + \gamma]\, P_{\underline{\theta}}\{i \in S(\underline{X})\}$. Clearly, the latter is not an expectation. It should be treated as an inner product, where its Schur-type monotonicity is relevant for performance considerations.

3. Bayes Risks and new Procedures

Let $X_1,\ldots,X_k$ be the sample means, based on independent samples of sizes $n_1,\ldots,n_k$, from k normal populations $P_1,\ldots,P_k$, respectively. It is assumed that $X_i \sim N(\theta_i, p_i^{-1})$, where $p_i = n_i / \sigma_i^2$ denotes the precision of the sample mean X_i, and that $\sigma_i^2 > 0$ is known, $i = 1,\ldots,k$. The performance of a selection rule S is measured by its expected loss, i.e. by its frequentist risk $R(\underline{\theta},S) = E_{\underline{\theta}}(L(\underline{\theta},S(\underline{X}))$, at parameter configurations $\underline{\theta} \in \Re^k$.

For admissibility considerations of subset selection rules S, the present framework will now be extended to the Bayesian approach. Starting with the pioneering work by Dunnett (1960), Bayesian selection procedures have been studied at numerous occasions in the literature. An

overview of these results can be found in Miescke (1997). From now on, it is assumed that the parameters $\underline{\Theta} = (\Theta_1,...,\Theta_k)$, say, are a priori random variables which follow a known prior density $\pi(\underline{\theta})$, $\underline{\theta} \in \Re^k$. More specifically, it will be assumed below that a priori, $\Theta_i \sim N(\mu_i, \nu_i^{-1})$, $i = 1,...,k$, where the prior mean $\mu_i \in \Re$ and the precision $\nu_i > 0$, $i = 1,...,k$, are known, and where $\Theta_1,...,\Theta_k$ are independent. A posteriori, given $\underline{X} = \underline{x}$, $\Theta_i \sim N(\mu_i(x_i), \tau_i^{-1})$, with $\mu_i(x_i) = (p_i x_i + \nu_i \mu_i)/(p_i + \nu_i)$ and $\tau_i = p_i + \nu_i$, $i = 1,...,k$, and $\Theta_1,...,\Theta_k$ are independent. Moreover, marginally, $X_i \sim N(\mu_i, p_i^{-1} + \nu_i^{-1})$, $i = 1,...,k$, and are independent.

Let $\Omega = \{\underline{\theta} \mid \theta_{[k-1]} < \theta_{[k]}, \underline{\theta} \in \Re^k\}$ be that part of the parameter space where the largest of the k parameters is unique. Since $P\{\underline{\Theta} \in \Omega\} = 1$, the parameter space can be replaced by Ω, which thus will be done now.

First, the *"(0,a,a-1) loss"* will be considered. Since all three loss functions in (2) are identical on Ω, for convenience, $L(\underline{\theta}, s) = L_1(\underline{\theta}, s)$, $\underline{\theta} \in \Omega$, will be used in the sequel. It should be noted that $a - 1 \le L(\underline{\theta}, s) \le (k-1)a$ holds, and that the two bounds can be attained. On Ω, the frequentist risk $R(\underline{\theta}, S) = E_{\underline{\theta}}(L(\underline{\theta}, S(\underline{X}))$ of a subset selection rule S under loss L is given by (4), and it inherits the two bounds from L. Likewise, the lower and the upper bound can be attained, as it can be seen from (4) by using a no-data rule that selects 1 or k-1, respectively, fixed populations.

The Bayes rules S^π for the normal prior density $\pi(\underline{\theta})$, $\underline{\theta} \in \Re^k$, under loss function L, minimize the posterior risk, i.e. the posterior expected loss, at every $\underline{X} = \underline{x} \in \Re^k$, and are

$$S^\pi(\underline{x}) = \begin{cases} \left\{ i \,\middle|\, P\left\{\Theta_i = \Theta_{[k]} \,\middle|\, \underline{X} = \underline{x}\right\} \ge a, \quad i = 1,...,k \right\}, \text{ if this set is not empty,} \\ \{i_0\} \text{ for any } i_0 \text{ with } P\left\{\Theta_{i_0} = \Theta_{[k]} \,\middle|\, \underline{X} = \underline{x}\right\} = \max\limits_{j=1,...,k} P\left\{\Theta_j = \Theta_{[k]} \,\middle|\, \underline{X} = \underline{x}\right\}, \text{ otherwise,} \end{cases} \quad (8)$$

with the option to drop elements in the first set for which equality occurs. This follows from the fact that the Bayes risk, which is the expected minimum of the $2^k - 1$ posterior risks that are associated with all non-empty subset selections $s(\underline{X})$, is equal to

$$r(\pi, S^{\pi}) = E\left(\min_{s(\underline{X}) \neq \emptyset} \sum_{i \in s(\underline{X})} \left[a - P\left\{ \Theta_i = \Theta_{[k]} \middle| \underline{X} \right\} \right] \right). \tag{9}$$

The Bayes risk of any subset selection procedure S, including S^{π}, can be represented by

$$r(\pi, S) = E\left(\sum_{i \in S(\underline{X})} \left[a - P\left\{ \Theta_i = \Theta_{[k]} \middle| \underline{X} \right\} \right] \right). \tag{10}$$

The lower bound $a - 1$ and the upper bound $(k - 1)a$ on the frequentist risk are inherited by the Bayes risk, but may not be attained by the latter.

A classical approach to verify admissibility is the method by Blyth (1951). Similar as in Miescke and Park (1997), it will be shown that all assumptions, including **(a)**, **(b)**, and **(c)**, of *Theorem 13* in Berger (1985, p. 547), which provides admissibility, can be met. Although Ω is not convex, it differs from $\Re^k$ only by a Lebesgue null set, and thus is sufficient for the proof of that theorem to remain valid.

In the present setting, the frequentist risk of every subset selection procedure S is a continuous function of $\underline{\theta} \in \Omega$. This follows from the fact that under loss L, the risk of S at any $\underline{\theta} \in \Omega$, with $\theta_{i^*(\underline{\theta})} = \theta_{[k]}$, which is given by (4), can be written as

$$R(\underline{\theta}, S) = a \sum_{i=1}^{k} P_{\underline{\theta}}\left\{ i \in S(\underline{X}) \right\} - P_{\underline{\theta}}\left\{ i^*(\underline{\theta}) \in S(\underline{X}) \right\}, \tag{11}$$

where $P_{\underline{\theta}}\left\{ r \in S(\underline{X}) \right\}$ is continuous on Ω, $r = 1, \ldots, k$, since the underlying distribution is a regular k-parameter exponential family. It follows that all Bayes rules for the prior density $\pi(\underline{\theta})$ are admissible on Ω under loss L in the class of non-randomized subset selection rules.

Moreover, a subset selection rule S which is admissible on Ω must be admissible on $\Re^k$, which can be seen easily from (3), and the fact that $P_{\underline{\theta}}\{r \in S(\underline{X})\}$ is also continuous on $\Re^k$, $r = 1,\ldots,k$.

The set $N = \{\underline{x} \mid x_{[k-1]} = x_{[k]}\}$ is a Lebesgue null set in $\Re^k$, which has no effect on the frequentist and Bayes risks considered below. Therefore, the sample space can also be replaced by Ω, as long as Bayes risks for prior density $\pi(\underline{\theta})$ on $\Re^k$ are under concern. This proves to be convenient and will thus be done in the sequel. For $n = 1,2,\ldots$, let π_n be the prior density $\pi(\underline{\theta})$ with the choices of $\mu_i = 0$ and $\nu_i^{-1} = n/p_i$, $i = 1,\ldots,k$. Furthermore, let π_n^* be the improper prior density given by $\pi_n^*(\underline{\theta}) = n^{k/2}\pi_n(\underline{\theta})$, i.e.

$$\pi_n^*(\underline{\theta}) = (2\pi)^{-k/2} \prod_{i=1}^{k} \exp\left(-(2n)^{-1}\theta_i^2\right), \quad \underline{\theta} \in \Re^k . \tag{12}$$

(a) The Bayes risks $r(\pi_n^*, S)$ of any subset selection procedure S, and thus also $r(\pi_n^*, S^{\pi_n})$ of S^{π_n}, which is the Bayes procedure with respect to π_n (as well as the generalized Bayes procedure with respect to π_n^*), are finite for all $n = 1,2,\ldots$, since the loss function L, given by L_1 in (2), is bounded by $a - 1 \le L(\underline{\theta}, s) \le k(a-1)$.

(b) For any convex set $K \subset \Re^k$ that is non-degenerate, i.e. that has a positive Lebesgue measure, there exists a $Q > 0$ and an integer M such that for $n \ge M$,

$$\int_K \pi_n^*(\underline{\theta})\, d\underline{\theta} \ge Q . \tag{13}$$

This follows from the fact that $\pi_1^* < \pi_n^*$ for every integer $n > 1$.

(c) To prove that a subset selection procedure S is admissible on Ω under loss L in the class of all non-randomized subset selection rules, it is sufficient to show that

$$\lim_{n\to\infty}\left[r(\pi_n^*,S)-r(\pi_n^*,S^{\pi_n})\right]=\lim_{n\to\infty}n^{k/2}\left[r(\pi_n,S)-r(\pi_n,S^{\pi_n})\right]=0. \tag{14}$$

The Bayes risk of the Bayes subset selection procedure S^{π_n} for prior π_n, $n=1,2,\ldots$, will now be examined in details. After rewriting (9) in terms of integrals, one has for $n=1,2,\ldots$,

$$r(\pi_n,S^{\pi_n})=\int_\Omega \min_{s(\underline{x})\neq\varnothing}\sum_{i\in s(\underline{x})}\left[a-\int_{\Re}\prod_{j\neq i}\Phi\Big(p_j^{1/2}\Big[[n/(n+1)]^{1/2}(x_i-x_j)+p_i^{-1/2}z\Big]\Big)\varphi(z)dz\right] \tag{15}$$
$$\times\prod_{r=1}^{k}[p_r/(n+1)]^{1/2}\varphi\Big([p_r/(n+1)]^{1/2}x_r\Big)\,dx_r,$$

The minimum in (15) is taken pointwise, at every $\underline{x}\in\Omega$, over the 2^k-1 non-empty subsets of $\{1,\ldots,k\}$. A change of variables $x_r=\sqrt{n+1}\,z_r$, $r=1,\ldots,k$, leads to the alternative representation

$$r(\pi_n,S^{\pi_n})=\int_\Omega \min_{s(\underline{z})\neq\varnothing}\sum_{i\in s(\underline{z})}\left[a-\int_{\Re}\prod_{j\neq i}\Phi\Big(p_j^{1/2}\Big[n^{1/2}(z_i-z_j)+p_i^{-1/2}z\Big]\Big)\varphi(z)dz\right] \tag{16}$$
$$\times\prod_{r=1}^{k}p_r^{1/2}\varphi(p_r^{1/2}z_r)\,dz_r.$$

The Bayes risk of any subset selection procedure S can be seen in the same way to have, analogously to (15) and (16), the two representations for $n=1,2,\ldots$

$$r(\pi_n,S)=\int_\Omega\sum_{i\in S(\underline{x})}\left[a-\int_{\Re}\prod_{j\neq i}\Phi\Big(p_j^{1/2}\Big[[n/(n+1)]^{1/2}(x_i-x_j)+p_i^{-1/2}z\Big]\Big)\varphi(z)dz\right] \tag{17}$$
$$\times\prod_{r=1}^{k}[p_r/(n+1)]^{1/2}\varphi\Big([p_r/(n+1)]^{1/2}x_r\Big)\,dx_r$$
$$=\int_\Omega\sum_{i\in S((n+1)^{1/2}\underline{z})}\left[a-\int_{\Re}\prod_{j\neq i}\Phi\Big(p_j^{1/2}\Big[n^{1/2}(z_i-z_j)+p_i^{-1/2}z\Big]\Big)\varphi(z)dz\right]\prod_{r=1}^{k}p_r^{1/2}\varphi\Big(p_r^{1/2}z_r\Big)\,dz_r.$$

As in (15) and (16), $\underline{x}$ plays the role of the k sample means $\underline{X}=\underline{x}\in\Omega$, and $\underline{z}=(n+1)^{-1/2}\underline{x}$.

An interesting subset selection rule is the generalized Bayes rule S^{Gen} for the Lebesgue measure used as non-informative prior on $\Re^k$. It selects, at $\underline{x}\in\Re^k$, any subset which achieves

$$r(\underline{x}) := \min_{s(\underline{z})\neq\emptyset} \sum_{i\in s(\underline{z})} \left[a - \int\prod_{\Re\ j\neq i} \Phi\left(p_j^{1/2}\left[x_i - x_j + p_i^{-1/2} z \right] \right) \varphi(z)dz \right]. \tag{18}$$

Lemma 2. *For every subset selection procedure* S, *and for every* $n = 1,2,\ldots,$

$$0 \le r(\pi_n, S) - (a-1) \tag{19}$$
$$\le \int_{\Omega} \sum_{i\in S((n+1)^{1/2}\underline{z})} \left[1 - \int\prod_{\Re\ j\neq i} \Phi\left(p_j^{1/2}\left[n^{1/2}(z_i - z_j) + p_i^{-1/2} z \right] \right) \varphi(z)dz \right] \prod_{r=1}^{k} p_r^{1/2} \varphi\left(p_r^{1/2} z_r \right) dz_r .$$

Proof: It has been shown already that $r(\pi_n, S) \ge a - 1$ holds. As to the second inequality, there is at least one term in the (first and) second sum of (17). Thus, the sum plus $(1-a)$ is less or equal than the sum of each term plus $(1-a)$.

The limiting behavior of the Bayes risks of the subset selection rules considered above is as follows.

Theorem 1. $\lim_{n\to\infty} r(\pi_n, S^{Gen}) = \lim_{n\to\infty} r(\pi_n, S_G) = \lim_{n\to\infty} r(\pi_n, S^{\pi_n}) = a - 1.$

Proof: Let $\underline{z} \in \Omega$ be fixed. The inner integral of (19) converges, as n tends to infinity, to 1 if $i = i^*(\underline{z})$ with $z_{i^*(\underline{y})} = z_{[k]}$, and it converges to 0 otherwise. For S^{Gen}, the minimum in (18), by using $r(\underline{x}) = r((n+1)^{1/2}\underline{z})$, occurs, for sufficiently large n, at the subset $s^*(\underline{z}) = \{i^*(\underline{z})\}$. Thus, by Lebesgue's bounded convergence theorem, one can verify that for S^{Gen} the Bayes risks in (19) tends to $a - 1$ as n tends to infinity. Similarly, one can verify the limit for any subset selection rule of the type S_G, as defined below of (1), using $S_G(\underline{x}) = S_G((n+1)^{1/2}\underline{z})$. The third limit follows from the fact that $a - 1 \le r(\pi_n, S^{\pi_n}) \le r(\pi_n, S^{Gen})$ holds for all n.

The limiting behavior of $n^{k/2}\left[r(\pi_n, S^{\pi_n}) - (a-1) \right]$, as n tends to infinity, remains unresolved. In view of (16), at $\underline{z} \in \Omega$, and for $i^*(\underline{z})$ given by $z_{i^*(\underline{z})} = z_{[k]}$,

$$\lim_{n\to\infty} n^{k/2} \int_{\Omega} I_{S^{\pi_n}(\underline{z})}(i^*(\underline{z}))\left[1 - \int_{\Re} \prod_{j\neq i^*(\underline{z})} \Phi\left(p_j^{1/2}\left[n^{1/2}(z_{i^*(\underline{z})} - z_j) + p_{i^*(\underline{z})}^{-1/2} z\right]\right)\varphi(z)dz\right] \tag{20}$$

$$\times \prod_{r=1}^{k} p_r^{1/2}\varphi(p_r^{1/2} z_r)\, dz_r = \infty .$$

This can be seen by changing variables in (20) with $v_r = n^{k/2} z_r$, $r = 1,\ldots,k$, and then considering an area of $\underline{v} \in \Omega$ where the differences of the coordinates are smaller than a suitable bound. On the other hand, for $i \neq i^*(\underline{z})$, the limit of the expression in (20), with $i^*(\underline{z})$ replaced by i, and with 1 replaced by a , is only known to be less or equal to zero. This difficulty makes in unfeasible to determine if (14) does or does not hold for S^{Gen} or S_G.

For every $n = 1,2,\ldots$, the Bayes rule S^{π_n} for prior π_n is admissible and thus has the optimum properties in term of the probability of a correct selection and the expected subset size given by Lemma 1 on $\Re^k$. From a practical point of view, the difference between S^{Gen} and S^{π_n} becomes negligible for large n, and therefore S^{Gen} appears to be suitable for practical use.

To establish this new subset selection procedure S^{Gen} at the P^*-condition, one has to compromise on the value of a in the loss function and use that value of a for which P^* is equal to the minimum of the probability of a correct selection. To determine the latter, let

$$q_i(\underline{x}) = \int_{\Re} \prod_{j\neq i} \Phi\left(p_j^{1/2}\left[x_i - x_j + p_i^{-1/2} z\right]\right)\varphi(z)dz, \quad i = 1,\ldots,k, \quad \underline{x} \in \Re^k . \tag{21}$$

From (18) it follows that at every $\underline{x} \in \Re^k$,

$$S^{Gen}(\underline{x}) = \begin{cases} \left\{ i \,\middle|\, q_i(\underline{x}) \geq a, \quad i = 1,\ldots,k \right\}, & \text{if this set is not empty,} \\ \{i_0\}, & \text{for any } i_0 \text{ with } q_{i_0}(\underline{x}) = \max\limits_{j=1,\ldots,k} q_j(\underline{x}), \text{ otherwise,} \end{cases} \tag{22}$$

with the option to drop elements in the first set for which equality occurs. Using the representation $X_r = \theta_r + p_r^{-1/2} N_r$, $r = 1,...,k$, where $N_1,...,N_k$ are generic i.i.d. standard normal random variables, the probability of a correct selection of S^{Gen} at $\underline{\theta} \in \Omega$ with $\theta_{i^*(\underline{\theta})} = \theta_{[k]}$ is

$$P\Bigg\{ \int_{\Re} \prod_{j \neq i^*(\underline{\theta})} \Phi\Big(p_j^{1/2}\Big[\theta_{[k]} - \theta_j + p_{i^*(\underline{\theta})}^{-1/2} N_{i^*(\underline{\theta})} - p_j^{-1/2} N_j + p_{i^*(\underline{\theta})}^{-1/2} z\Big]\Big) \varphi(z)dz \geq a, \text{ or} \tag{23}$$

$$\int_{\Re} \prod_{j \neq i^*(\underline{\theta})} \Phi\Big(p_j^{1/2}\Big[\theta_{[k]} - \theta_j + p_{i^*(\underline{\theta})}^{-1/2} N_{i^*(\underline{\theta})} - p_j^{-1/2} N_j + p_{i^*(\underline{\theta})}^{-1/2} z\Big]\Big) \varphi(z)dz$$

$$\geq \int_{\Re} \prod_{j \neq r} \Phi\Big(p_j^{1/2}\Big[\theta_r - \theta_j + p_r^{-1/2} N_r - p_j^{-1/2} N_j + p_r^{-1/2} z\Big]\Big) \varphi(z)dz, \quad \text{for all } r \neq i^*(\underline{\theta}) \Bigg\}$$

It has been shown in Gupta and Miescke (1988) that for $x_1 = x_2 = ... = x_k$, the maximum of the k values in (21) occurs at $\{i \mid p_i = p_{[1]}, i = 1,...,k\}$. Thus, if $x_1,...,x_k$ are sufficiently close together, the maximum of the k values in (21) may not occur at $i = i^*(\underline{x})$. However, for $p_1 = p_2 = ... = p_k$, that maximum always occurs at $i = i^*(\underline{x})$, and here S^{Gen} can be established at the P^*-condition with a value of a that depends only on k and P^*.

Theorem 2. *For* $p_1 = p_2 = ... = p_k = p$, say, S^{Gen} *satisfies the* P^**-condition iff* a *satisfies*

$$P\Bigg\{ N_k = N_{[k]}, \text{ or } \int_{\Re} \prod_{j=1}^{k-1} \Phi\Big(N_k - N_j + z\Big) \varphi(z)dz \geq a \Bigg\} = P^*. \tag{24}$$

Proof: From (21) - (23), one can see that for $p_1 = p_2 = ... = p_k = p$, the probability of a correct selection under S^{Gen} at $\underline{\theta} \in \Omega$ with $\theta_k = \theta_{[k]}$ is

$$P\Big\{ \theta_{[k]} + p^{-1/2} N_k \geq \theta_j + p^{-1/2} N_j, \text{ for all } j < k, \tag{25}$$

$$\text{or } \int_{\Re} \prod_{j=1}^{k-1} \Phi\Big(p^{1/2}\Big[\theta_{[k]} - \theta_j\Big] + N_k - N_j + z\Big) \varphi(z)dz \geq a \Big\},$$

which is decreasing in $\theta_{[k]}$ and increasing in $\theta_{[1]},\ldots,\theta_{[k-1]}$. The minimum of (25) thus occurs at $\theta_1 = \theta_2 = \ldots = \theta_k$, where it is equal to (24).

For any predetermined P^*, the value of a for which (24) holds has to be determined on a computer with numerical integration or simulation. Likewise, comparisons of the expected subset sizes of S^{Gen} and another subset selection procedure, such as S_G, both meeting the P^*-condition, has to be done in this way.

In the second part of this section, considerations similar to those above will now be made for the *"linear loss"* (6), but in a more concise manner. The loss function $L^\bullet$ has $-\Delta$ as a lower bound, which can be achieved. This bound is inherited by the frequentist risk and the Bayes risk, and can be achieved by the former.

The Bayes subset selection rules $S^\pi_\bullet$ for the normal prior density $\pi(\underline{\theta})$, $\underline{\theta} \in \Re^k$, under loss function $L^\bullet$, minimizes the posterior risk at every $\underline{X} = \underline{x} \in \Re^k$, and are

$$S^\pi_\bullet(\underline{x}) = \begin{cases} \left\{ i \,\middle|\, E\left\{\Theta_i \,\middle|\, \underline{X} = \underline{x}\right\} \geq E\left\{\Theta_{[k]} \,\middle|\, \underline{X} = \underline{x}\right\} - \Delta,\ i = 1,\ldots,k \right\}, & \text{if this set is not empty,} \\ \{i_0\} \text{ for any } i_0 \text{ with } E\left\{\Theta_{i_0} \,\middle|\, \underline{X} = \underline{x}\right\} = \max\limits_{j=1,\ldots,k} E\left\{\Theta_j \,\middle|\, \underline{X} = \underline{x}\right\}, & \text{otherwise,} \end{cases} \tag{26}$$

with the option to drop elements in the first set from which equality occurs. This follows from the fact that the Bayes risk of the Bayes subset selection rules $S^\pi_\bullet$ is given by

$$r_\bullet(\pi, S^\pi_\bullet) = E\left(\min_{s(\underline{X}) \neq \emptyset} \sum_{i \in s(\underline{X})} \left[E\left\{\Theta_{[k]} \,\middle|\, \underline{X}\right\} - E\left\{\Theta_i \,\middle|\, \underline{X}\right\} - \Delta \right] \right), \tag{27}$$

and the Bayes risk of any subset selection procedure S is given by

$$r_\bullet(\pi, S) = E\left(\sum_{i \in S(\underline{X})} \left[E\left\{\Theta_{[k]} \,\middle|\, \underline{X}\right\} - E\left\{\Theta_i \,\middle|\, \underline{X}\right\} - \Delta \right] \right). \tag{28}$$

The Bayes risk of the Bayes subset selection procedure $S_\bullet^{\pi_n}$ under loss L° for prior π_n, $n = 1,2,\ldots$, which are given above of (12), is

$$r_\bullet(\pi_n, S_\bullet^{\pi_n}) \tag{29}$$
$$= \int_\Omega \min_{s(\underline{x})\neq\emptyset} \sum_{i\in s(\underline{x})} \left[E\left(\max_{j=1,\ldots,k} \left\{ [n/(n+1)]x_j + [n/(n+1)]^{1/2} p_j^{-1/2} N_j \right\} \right) - [n/(n+1)]x_i - \Delta \right]$$
$$\times \prod_{r=1}^{k} [p_r/(n+1)]^{1/2} \varphi\left([p_r/(n+1)]^{1/2} x_r \right) dx_r,$$

where $N_1,\ldots,N_k$ are generic i.i.d. standard normal random variables. The minimum in (29) is taken pointwise, at every $\underline{x} \in \Omega$, over all $2^k - 1$ non-empty subsets of $\{1,\ldots,k\}$. A change of variables $x_r = \sqrt{n+1}\, z_r$ $r = 1,\ldots,k$, leads to the alternative representation

$$r_\bullet(\pi_n, S_\bullet^{\pi_n}) = \int_\Omega \min_{s(\underline{z})\neq\emptyset} \sum_{i\in s(\underline{z})} \left[[n/(n+1)]^{1/2} \left\langle E\left(\max_{j=1,\ldots,k} \left\{ n^{1/2} z_j + p_j^{-1/2} N_j \right\} \right) - n^{1/2} z_i \right\rangle - \Delta \right] \tag{30}$$
$$\times \prod_{r=1}^{k} p_r^{1/2} \varphi\left(p_r^{1/2} z_r \right) dz_r .$$

The Bayes risk $r_\bullet(\pi_n, S)$ of any other subset selection procedure S can be derived from (29) by deleting the minimum and replacing $s(\underline{x})$ by $S(\underline{x})$. Likewise, it can be derived from (30) by deleting the minimum and replacing $s(\underline{z})$ by $S((n+1)^{1/2}\underline{z})$.

An interesting subset selection rule is the generalized Bayes rule $S_\bullet^{Gen}$ for the Lebesgue measure used as a non-informative prior on $\Re^k$. It selects, at $\underline{x} \in \Re^k$, any subset which achieves

$$r_\bullet(\underline{x}) := \min_{s(\underline{x})\neq\emptyset} \sum_{i\in s(\underline{x})} \left[E\left(\max_{j=1,\ldots,k} \left\{ x_j + p_j^{-1/2} N_j \right\} \right) - x_i - \Delta \right]. \tag{31}$$

The limiting behavior of the Bayes risks of the subset selection rules considered above is as follows.

Theorem 3. $\lim_{n\to\infty} r_\bullet(\pi_n, S_\bullet^{Gen}) = \lim_{n\to\infty} r_\bullet(\pi_n, S_G) = \lim_{n\to\infty} r_\bullet(\pi_n, S_\bullet^{\pi_n}) = -\Delta.$

Proof: Let $\underline{x} \in \Omega$ be fixed. Obviously, $i^*(\underline{x})$ is element of S_G, $S_\bullet^{Gen}$, and $S_\bullet^{\pi_n}$ for all n. That part of the sum in the respective Bayes risks which is associated with a selection of $i^*(\underline{x})$, in the setting of (30) with $\underline{z} = (n+1)^{-1/2}\underline{x}$, tends to $-\Delta$. This follows from

$$-\Delta \le [n/(n+1)]^{1/2} E\left(\max_{j=1,\dots,k}\left\{n^{1/2}(z_j - z_{[k]}) + p_j^{-1/2}N_j\right\}\right) - \Delta \le E\left(\max_{j=1,\dots,k}\left\{p_j^{-1/2}N_j\right\}\right) - \Delta, \quad (32)$$

and the fact that the term in the term in the middle of (32) tends to 0 as n tends to infinity.

Let now $i \neq i^*(\underline{x})$. Since for $i \in S_G(\underline{x})$, $x_{[k]} - x_i \le \delta$, i.e. $(n+1)^{1/2}(z_{[k]} - z_i) \le \delta$, for some $\delta > 0$, that part of the sum in the Bayes risk of S_G which is associated with a selection of i is bounded by

$$\begin{aligned} -\Delta &\le I_{S_G((n+1)^{1/2}\underline{z})}(i)\left[[n/(n+1)]^{1/2} E\left(\max_{j=1,\dots,k}\left\{n^{1/2}(z_j - z_i) + p_j^{-1/2}N_j\right\}\right) - \Delta\right] \quad (33)\\ &\le \delta + E\left(\max_{j=1,\dots,k}\left\{p_j^{-1/2}N_j\right\}\right) - \Delta, \end{aligned}$$

where the indicator function in (33) is 0 for sufficiently large n. Thus by Lebesgue's bounded convergence theorem, the second limit in the theorem holds true. The third follows from $-\Delta \le r_\bullet(\pi_n, S_\bullet^{\pi_n}) \le r_\bullet(\pi_n, S_G)$, $n = 1,2,\dots$

For any $i \neq i^*(\underline{x})$ with $i \in S_\bullet^{Gen}(\underline{x})$, that part of the sum in the Bayes risk of $S_\bullet^{Gen}$ which is associated with a selection of i is bounded by

$$\begin{aligned} -\Delta &\le I_{S_\bullet^{Gen}((n+1)^{1/2}\underline{z})}(i)\left[[n/(n+1)]^{1/2} E\left(\max_{j=1,\dots,k}\left\{n^{1/2}(z_j - z_i) + p_j^{-1/2}N_j\right\}\right) - \Delta\right] \quad (34)\\ &\le I_{S_\bullet^{Gen}((n+1)^{1/2}\underline{z})}(i)\left[E\left(\max_{j=1,\dots,k}\left\{(n+1)^{1/2}(z_j - z_i) + [(n+1)/n]^{1/2}p_j^{-1/2}N_j\right\}\right) - \Delta\right]\\ &\le \left\langle[(n+1)/n]^{1/2} - 1\right\rangle E\left(\max_{j=1,\dots,k}\left\{p_j^{-1/2}N_j\right\}\right) \le E\left(\max_{j=1,\dots,k}\left\{p_j^{-1/2}N_j\right\}\right). \end{aligned}$$

The second inequality in (34) is established by replacing a factor $n/(n+1)$ of the expectation by one. The third follows from the fact that, according to (31),

$$E\left(\max_{j=1,\ldots,k}\left\{(n+1)^{1/2}(z_j - z_i) + p_j^{-1/2}N_j\right\}\right) \le \Delta \quad \text{for} \quad i \in S_\bullet^{Gen}((n+1)^{1/2}(\underline{z})). \tag{35}$$

This completes the proof of the theorem.

Similar as with the *"(0,a,a-1) loss"*, the limiting behavior of $n^{k/2}\left[r_o(\pi_n, S_o^{\pi_n}) + \Delta\right]$, as n tends to infinity, remains unresolved. This difficulty makes in unfeasible to determine if (14) does or does not hold for $S_\bullet^{Gen}$ or S_G.

From (7) it can be seen that the frequentist risk of every Bayes subset selection procedure is continuous on $\Re^k$, and thus all Bayes rules for the prior densities $\pi(\underline{\theta})$ and $\pi_n(\underline{\theta})$ are admissible on $\Re^k$ in the class of non-randomized subset selection rules. The arguments are the same that were used below of (11). From a practical point of view, the difference between $S_\bullet^{Gen}$ and $S_\bullet^{\pi_n}$ becomes negligible for large n, and thus $S_\bullet^{Gen}$ appears to be suitable for practical use.

To establish this new subset selection procedure $S_\bullet^{Gen}$ at the P^*-condition, one has to compromise on the value of Δ in the loss function L^* and use that value of Δ for which P^* is equal to the minimum of the probability of a correct selection. To determine the latter, let

$$\rho_i(\underline{x}) = x_i - E\left(\max_{j=1,\ldots,k}\left\{x_j + p_j^{-1/2}N_j\right\}\right), \quad i = 1,\ldots,k, \quad \underline{x} \in \Re^k. \tag{36}$$

From (31) it follows that at every $\underline{x} \in \Re^k$,

$$S_\bullet^{Gen}(\underline{x}) = \begin{cases} \left\{ i \,\middle|\, \rho_i(\underline{x}) \ge -\Delta, \quad i = 1,\ldots,k \right\}, & \text{if this set is not empty,} \\ \{i_0\}, & \text{for any } i_0 \text{ with } x_{i_0} = x_{[k]}, \text{ otherwise,} \end{cases} \tag{37}$$

with the option to drop elements in the first set for which equality occurs. The probability of a correct selection of $S_{\bullet}^{Gen}$ at $\underline{\theta} \in \Omega$ with $\theta_{i^*(\underline{\theta})} = \theta_{[k]}$ is

$$P^{\underline{Z}}\Big\{ \theta_{[k]} + p_{i^*(\underline{\theta})}^{-1/2} Z_{i^*(\underline{\theta})} - E^{\underline{N}}\Big(\max_{j=1,\ldots,k}\Big\{\theta_j + p_j^{-1/2} Z_j + p_j^{-1/2} N_j\Big\}\Big) \geq -\Delta, \quad \text{or} \qquad (38)$$
$$\theta_{[k]} + p_{i^*(\underline{\theta})}^{-1/2} Z_{i^*(\underline{\theta})} \geq \theta_r + p_r^{-1/2} Z_r \text{ , for all } r \neq i^*(\underline{\theta})\Big\},$$

where $Z_1,\ldots,Z_k$ are i.i.d. generic standard normal random variables which are independent of $N_1,\ldots,N_k$, and where the superscripts indicate the random variables involved. $S_{\bullet}^{Gen}$ can be established at the P^*-condition with a value of Δ that depends on $p_1,\ldots,p_k$, k, and P^*.

Theorem 4. $S_{\bullet}^{Gen}$ *satisfies the* P^**-condition iff* Δ *satisfies*

$$\min_{i=1,\ldots,k} P^{\underline{Z}}\Big\{ p_i^{-1/2} Z_i = \max_{j=1,\ldots,k}\Big\{p_j^{-1/2} Z_j\Big\} \text{ , or } p_i^{-1/2} Z_i - E^{\underline{N}}\Big(\max_{j=1,\ldots,k}\Big\{p_j^{-1/2}(Z_j + N_j)\Big\}\Big) \geq -\Delta \Big\} = P^*. \qquad (39)$$

Proof: The assertion follows from the fact that the probability in (38) is increasing in $\theta_{[k]}$ and decreasing in $\theta_{[1]},\ldots,\theta_{[k-1]}$.

It is interesting to note that for $p_1 = p_2 = \ldots = p_k = p$, say, (39) simplifies to

$$P^{\underline{Z}}\Big\{ Z_k = Z_{[k]} \text{ , or } Z_k - E^{\underline{N}}\Big(\max_{j=1,\ldots,k}\Big\{Z_j + N_j\Big\}\Big) \geq -p^{1/2}\Delta \Big\} = P^*. \qquad (40)$$

For any predetermined P^*, the value of Δ for which (39) or (40) holds has to be determined on a computer with numerical integration or simulation. Likewise, comparisons of the expected subset sizes of $S_{\bullet}^{Gen}$ and another subset selection procedure, such as S_G, both meeting the P^*-condition, has to be done in this way.

In conclusion, the extension of the admissibility results from non-randomized to randomized subset selection rules will be described briefly. A randomized subset selection rule

S^*, say, can be represented by $S^*(\underline{x},\underline{u}) = \left\{ i \,\middle|\, u_i \le p_{S^*,i}(\underline{x}), i = 1,\ldots,k \right\}$, where at every $\underline{x} \in \Re^k$, $p_{S^*,r}(\underline{x})$ denotes the probability of including population P_r into the subset, $r = 1,\ldots,k$, and where $U_1,\ldots,U_k$ are generic i.i.d. random variables, each uniformly distributed on [0,1], which are independent of $\underline{X}$. For such a randomized subset selection rule $S^*(\underline{X},\underline{U})$, one has

$$P_{\underline{\theta}}\{r \in S^*(\underline{X},\underline{U})\} = P_{\underline{\theta}}\left\{U_r \le p_{S^*,r}(\underline{X})\right\} = E_{\underline{\theta}}\left(p_{S^*,r}(\underline{X})\right), \quad \underline{\theta} \in \Re^k, r = 1,\ldots,k. \tag{41}$$

The risk function of a randomized subset selection rule S^* under loss function L is given by

$$R(\underline{\theta},S^*) = E_{\underline{\theta}}(L(\underline{\theta},S^*(\underline{X},\underline{U}))) = a\sum_{i=1}^{n} E_{\underline{\theta}}\left(p_{S^*,i}(\underline{X})\right) - \sum_{j \in A(\underline{\theta})} E_{\underline{\theta}}\left(p_{S^*,j}(\underline{X})\right), \quad \underline{\theta} \in \Re^k, \tag{42}$$

where $A(\underline{\theta}) = \{j | \theta_j = \theta_{[k]}\}$. By the same arguments that have been used before, for every randomized subset selection rule S^*, $R(\underline{\theta},S^*)$ is continuous on Ω. Every S^* that is admissible on Ω under L within the class of all randomized subset selection rules must also be admissible on $\Re^k$, since for any randomized subset selection rule S^*, $E_{\underline{\theta}}(p_{S^*,r}(\underline{X}))$ is continuous at every $\underline{\theta} \in \Re^k$, $r = 1,\ldots,k$. For loss function L^* similar arguments apply, but are omitted for brevity.

REFERENCES

Bechhofer, R.E., Santner, T.J., and Goldsman, D.M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. John Wiley, New York.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. Springer, N.Y.

Berger, R.L. and Gupta, S.S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. *Scand. J. Statist.* 7, 21-26.

Blyth, C.R. (1951). On minimax statistical decision procedures and their admissibility.

Ann. Math. Statist. **22**, 22-42.

Bratcher, T.L. and Bhalla, P. (1974). On the properties of an optimal selection procedure. *Commun. Statist.* **3**, 191-196.

Chen, H.J. and Dudewicz, E.J. (1973). Estimation of ordered parameters and subset selection. Division of Statistics, Ohio State University, *Technical Report.*

Deely, J. and Gupta, S.S. (1968). On the properties of subset selection procedures. *Sankhā* **A-30**, 37-50.

Deely J. and Johnson, W. (1997). Normal means revisited. Department of Statistics, Purdue University, West Lafayette, IN, *Technical Report* No. 97-01.

Dunnett, C.W. (1960). On selecting the largest of k normal population means. (With discussion) *J. Royal Statist. Soc.* **B-22**, 1-40.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Institute of Statistics, University of North Carolina, Chapel Hill, NC, *Mimeo Series* No. 150.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7**, 225-245.

Gupta, S.S. and Hsu, J.C. (1977). On the monotonicity of Bayes subset selection procedures. In: *Proceedings of the 41st Session of the ISI*, **Book 4**, 208-211.

Gupta, S.S. and Hsu, J.C. (1978). On the performance of some subset selection procedures. *Commun. Statist.-Simula.Computa.* **B-7**, 561-591.

Gupta, S.S. and Huang, D.Y. (1976). Subset selection procedures for the means and variances of normal populations: Unequal sample size case. *Sankhā* **B-38**, 112-128.

Gupta, S.S. and Huang, W.T. (1974). A note on selecting a subset of normal populations with unequal sample sizes. *Sankhā* **A-36**, 389-396.

Gupta, S.S. and Miescke, K.J. (1981). Optimality of subset selection procedures for ranking means of three normal populations. *Sankhā* **B-43**, 1-17.

Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. John Wiley, New York .

Gupta, S.S. and Wong, W.Y. (1976). Subset selection procedures for the means of normal populations with unequal variances: Unequal sample size case. Department of Statistics, Purdue University, *Technical Report* No. 473.

Miescke, K.J. (1979). Bayesian subset selection for additive and linear loss functions. *Commun. Statist.* **A-8**, 1205-1226.

Miescke, K.J. (1997). Bayesian sampling designs for selection procedures. In: *Multivariate, Design, and Sampling*, S. Ghosh ed., M. Dekker, New York, 34 pp. in ms., to appear.

Miescke, K.J. and Park, H. (1997). On the natural selection rule under normality. Research manuscript, 16 pp. in ms, submitted for publication.

Miescke, K.J. and Rasch, D., eds. (1996). Special Issue on: "40 Years of Statistical Selection Theory", Part I and II. J. Statist. Plann. Inf. **54**, No. 2 and 3.

Seal, K.C. (1955). On a class of decision procedures for ranking means of normal populations. *Ann. Math. Statist.* **26**, 387-398.

Seal, K.C. (1957). An optimum decision rule for ranking means of normal populations. *Calcutta Statist. Assoc. Bull.* **7**, 131-150.